

Modelos de regressão aplicados na previsão da evasão escolar do ensino básico: uma revisão sistemática da literatura

Flávio L. Morais¹, Ana Melo², Mirele Moutinho¹, Roberta Fagundes²

¹Universidade de Pernambuco (UPE)
Campus Caruaru – PE – Brasil

²Escola Politécnica de Pernambuco - POLI/UPE.

{flavio.leandromorais, roberta.fagundes, mirele.moutinho}@upe.br

accm@ecomp.poli.br

Abstract. *The educational system faces several challenges, among them is the school dropout, usually caused by social inequality. Due to this problem, Machine Learning (ML) techniques help to predict the relationship between the dropout student and the characteristics that influenced their dropout. This paper presents a Systematic Literature Review to identify which ML techniques, especially regression models, help to detect factors that influence the dropout of elementary and middle school students. The results shows low use of regression models, but high use of others ML models, highlighting the decision tree. In addition, 14 factors were identified as responsible for the dropout, calling attention to the low income of the students.*

Resumo. *O sistema educacional enfrenta diversos desafios, entre eles destaca-se a evasão escolar, geralmente causada pela desigualdade social. Técnicas de Machine Learning (ML) auxiliam a prever a relação entre o estudante evadido e as características que influenciaram a sua evasão. Esse artigo apresenta uma Revisão Sistemática da Literatura com o objetivo de identificar quais técnicas de ML, especialmente modelos de regressão, auxiliam na detecção de fatores que influenciam a evasão escolar de estudantes do ensino básico. Os resultados apresentam baixo uso de modelos de regressão, mas um alto uso de outros modelos de ML, destacando-se a árvore de decisão. Além disso, identificou-se 14 fatores responsáveis pela evasão, evidenciando a baixa renda do estudante.*

1. Introdução

A educação básica é uma ferramenta democrática que permite o desenvolvimento social e intelectual de um indivíduo, possibilitando a quebra de barreiras socioeconômicas, como também, garantindo a oportunidade de alcançar um padrão mínimo de qualidade da aprendizagem. Nos últimos anos, a educação básica passou por mudanças que configuraram melhorias estruturais para as instituições de ensino [Calixto et al. 2017]. No entanto, dados apresentam que o Brasil ainda se encontra entre os países com o pior desempenho nas áreas básicas da educação, como ciência e matemática [OECD 2015].

Dentre os problemas que o processo educacional enfrenta pode-se citar a evasão escolar. Segundo o Programa das Nações Unidas para o Desenvolvimento (PNDU), atualmente 262 milhões de crianças e adolescentes evadem do ensino básico no mundo

[do Desenvolvimento Humano 2019]. Esse dado é decorrente de várias causas provocadas pela desigualdade social que atinge diretamente as pessoas mais vulneráveis. Diante disso, o Fundo de Emergência Internacional das Nações Unidas para a Infância (UNICEF), relata os fatores correlacionados à evasão do estudante, tais como: a falta de transporte escolar, gravidez na adolescência e renda familiar baixa [UNICEF et al. 2020].

Vale ressaltar que o Brasil é uma das cinco nações que lidera o ranking de evasão escolar no mundo. Segundo os dados da UNICEF, cerca de 620 mil estudantes brasileiros do ensino básico evadiram no ano de 2019. Por conseguinte, a evasão escolar desencadeia uma série de adversidades que afetam diretamente o indivíduo e a sociedade, seja nos contextos educacionais, sociais, financeiros ou governamentais. Portanto, um dos problemas ocasionados pela evasão escolar é a redução de oportunidades para ingressar no mercado de trabalho, devido à falta de qualificação. Além disso, a baixa remuneração torna-se um fator preditivo para o desenvolvimento socioeconômico.

Contudo, ações governamentais em conjunto com as instituições de ensino estão desenvolvendo e aplicando políticas públicas na pretensão de reduzir o índice de evasão escolar. Neste sentido, destaca-se o projeto de diretriz do Espírito Santo [SEDU et al. 2018] e o sistema de detecção prévia do estado de São Paulo [Lopes Filho and Silveira 2021]. Ambos propõem identificar precocemente os estudantes em vulnerabilidade de evasão, e, dessa forma, propor medidas e políticas públicas educacionais mais ativas. Assim, é importante identificar previamente quais estudantes estão em risco de evadir [Adelman et al. 2018].

Nesse contexto, a literatura apresenta atualmente a Mineração de Dados Educacionais (do inglês, *Educational Data Mining - EDM*), como um relevante recurso para a identificação das causas que influenciam na evasão escolar, permitindo a extração de informações subjulgadas em grandes bancos de dados. A EDM é um campo de pesquisa que busca descobrir padrões sobre estudantes e formas de aprendizagem [Júnior et al. 2019]. Assim, destaca-se as técnicas de Aprendizagem de Máquina (do inglês, *Machine Learning - ML*), a exemplo, classificação, regressão e clusterização, tornaram-se fundamentais para analisar dados e produzir informações preditivas na identificação de estudantes vulneráveis à evasão.

Nessa perspectiva, ML busca programar a máquina com dados existentes, para que possa desenvolver experiência para a resolução de um determinado problema [Alpaydin 2020]. Portanto, prever a relação entre o estudante evadido e as características que influenciaram a evasão, são tópicos abordados por pesquisadores que utilizam técnicas de ML. Nesse sentido, relata-se a aplicação dos modelos de regressão, que segundo [do Nascimento et al. 2018], são ferramentas aplicadas de forma extensiva, gerando conhecimento e servindo como base para soluções de problemas e desenvolvimento de mecanismos em apoio ao ensino e a tomada de decisões.

À vista disso, é perceptível a baixa adesão da utilização dos modelos de regressão aplicados no contexto educacional, principalmente como foco em estimar a relação entre fatores educacionais e a evasão escolar. Sendo assim, esse artigo apresenta uma Revisão Sistemática da Literatura (RSL), com o objetivo de identificar as vertentes da utilização dos modelos de regressão na detecção de fatores que influenciam a evasão escolar de estudantes do ensino básico. Para isso, o método proposto

por [Kitchenham and Charters 2007] foi adotado. Iniciando-se com a elaboração do protocolo, no qual foram definidas as questões de pesquisa, bases de dados, *string* de busca, os critérios de seleção (inclusão e exclusão), como também, critérios de qualidade para análise e extração de informação dos artigos selecionados.

2. Trabalhos Relacionados

Com o objetivo de analisar o contexto da evasão escolar através da utilização de métodos de ML, em especial com modelos de regressão, foram realizadas buscas manuais a procura de outras RSL em conferências e periódicos. Entretanto, poucos estudos semelhantes de aspecto/contexto teórico foram encontrados, sendo estes apresentados abaixo.

Nesse contexto, o trabalho de [Colpo et al. 2020] realizou uma RSL com a meta de analisar estudos publicados nos anais do Congresso Brasileiro de Informática na Educação (CBIE) no período de 2006 a 2019, sobre o uso das técnicas de EDM na previsão da evasão do estudante em todas as fases de ensino. Ao final, 23 estudos foram selecionados e analisados e, dentre os resultados, apresentaram a falta de envolvimento dos profissionais na busca de soluções contra a evasão escolar.

Porém, o trabalho de [Marques et al. 2019] realizou um mapeamento sistemático com o objetivo de identificar as ferramentas, técnicas e fatores que auxiliam na identificação dos atributos que influenciam a evasão no ensino básico ou superior. Ao final 14 estudos foram considerados, e dentre os resultados, identificou-se que a ferramenta Weka, as técnicas de classificação e os fatores relacionados às características individuais do estudante, são destaques na descoberta de atributos que causam a evasão escolar.

Diante disso, o presente estudo diferencia-se por realizar dois tipos de busca (manual e automática). Além disso, foi realizado na área específica de mineração de dados com foco na utilização de modelos de regressão para previsão da evasão escolar do ensino básico. Nessa conjectura, relata-se que modelos de regressão apresentam aspectos que não são comumente encontrados na área educacional. No entanto, possuem maior precisão e confiabilidade, apoiando a intervenção e direcionamento dos estudantes em risco de evasão escolar. Adicionalmente, um dos pontos diferenciais nesse estudo, é coletar quais ações foram elicitadas como forma de mitigar a evasão escolar no ensino básico.

3. Revisão Sistemática da Literatura

Segundo [Kitchenham and Charters 2007], uma RSL é um estudo secundário que identifica, qualifica e interpreta todas as evidências de pesquisas primárias associadas a um fenômeno de interesse. Deste modo, esse trabalho conduz uma RSL com o objetivo de identificar a utilização dos modelos de regressão na predição de fatores que influenciam a evasão escolar de estudantes do ensino básico. Para isso, o processo de condução da RSL utiliza três fases: planejamento, condução e relatório.

3.1. Planejamento

3.1.1. Perguntas de Pesquisa

A etapa inicial do planejamento contempla a elaboração de Perguntas de Pesquisa relacionadas ao contexto da RSL proposta e de todos os processos para a sua condução, visto que o seu objetivo é responder essas indagações [Kitchenham and Charters 2007]. Diante

disso, desenvolveu-se a Pergunta de pesquisa Principal (PP) e 3 Perguntas de pesquisa Secundárias (PS) elicitadas a seguir:

PP: *Quais métodos de regressão são utilizados para prever a evasão escolar no ensino básico?*

PS1: *Quais fatores influenciam na evasão escolar no ensino básico?*

PS2: *Quais ações são adotadas para mitigar a evasão escolar no ensino básico?*

PS3: *Quais técnicas de machine learning estão sendo utilizados para auxiliar a tomada de decisão relacionada à evasão escolar no ensino básico?*

3.1.2. Engenhos e String de Busca

A princípio, propôs-se conduzir buscas automáticas e manuais com o objetivo de encontrar Estudos Primários (EP) relacionados ao objetivo deste trabalho. Os engenhos de buscas foram escolhidos pelo reconhecimento científico e associação com a área de tecnologia da informação e educação, comumente apresentadas no contexto desta pesquisa. Já sobre a ótica da quantidade de engenhos utilizada, acredita-se que foi o suficiente para o estudo, além de apresentar uma quantidade superior aos trabalhos relacionados citados anteriormente na Seção 2. Os engenhos de busca utilizados nesse estudo foram: Simpósio Brasileiro de Informática na Educação (SBIE); Revista Novas Tecnologias na Educação (RENTE); *International Conference on Educational Data Mining* (EDM); *IEEE Digital Library*; *ACM Digital Library*; *Scopus*; *Springer Link*; e *Science Direct*.

Logo após, buscou-se definir a string de busca, a qual foi determinada por meio de palavras chaves baseadas na PP. Em seguida, apropriou-se dos sinônimos e, por fim, foram efetuadas as combinações das palavras chaves e seus sinônimos, permitindo realizar-se o teste de sensibilidade¹, gerando ao final a seguinte string de busca:

((“School Dropout” OR “School Drop Outs” OR “School Drop-out” OR “School Abandon” OR “School Abandonment”) AND (“Regression methods” OR “regression techniques” OR “Data mining”))

3.1.3. Critérios de Seleção

Os critérios de seleção identificam os EP que fornecem evidências relacionadas às PP, e estão divididos nas classes de inclusão e exclusão, como seguem apresentados abaixo:

Critérios de Inclusão: (CI1) Estudos primários que respondam uma ou mais perguntas de pesquisa; (CI2) Estudos primários que foram publicados entre os anos de 2015 e 2021; (CI3) Estudos primários escritos em língua Inglesa e Portuguesa (Brasil); (CI4) Estudos primários que foram publicados em congressos, conferências ou revistas; e (CI5) Estudos que possuam como público alvo estudantes do ensino básico.

Critérios de Exclusão: (CE1) Estudos que não possuam conteúdo relacionado com o objetivo da pesquisa; (CE2) Estudos que possuam como público alvo o ensino superior;

¹encurtador.com.br/jnuOQ

(CE3) Estudos secundários e terciários; (CE4) Estudos que foram retornados como resultado em mais de uma base de dados; e (CE5) Estudos que foram inacessíveis.

3.1.4. Critérios de Qualidade

A avaliação de qualidade dos EP selecionados na fase anterior torna-se importante para qualificá-los, permitindo resultados sintetizados mais confiantes. Este trabalho baseou-se no estudo de [Dyba et al. 2007], onde se desenvolveram 11 critérios de qualidade, os quais foram adaptados e apresentados na Tabela 1. Ademais, os critérios de qualidade são organizados em quatro tipos, sendo eles relacionados:

Relatório: à qualidade da lógica dos objetivos e do contexto do estudo;

Rigor: à avaliação da credibilidade dos métodos de estudo para garantir que os recursos fossem válidos e significativos;

Credibilidade: ao rigor dos métodos de pesquisa empregados para estabelecer a validade das ferramentas de coleta de dados e dos métodos de análise;

Relevância: à avaliação da relevância do estudo para a indústria de software em geral e a comunidade de pesquisa.

Neste processo, todos os EP foram lidos na íntegra e, ao final, uma pontuação era atribuída. As possíveis pontuações eram: [0] - O estudo não atende ao critério de qualidade; [0,5] - O estudo atende parcialmente ao critério de qualidade; e [1] - O estudo satisfaz totalmente o critério de qualidade. Além disso, seguindo as recomendações de [Lima et al. 2019], era necessário uma pontuação mínima de 6 pontos para que o EP fosse considerado nesta RSL, ou seja, era preciso atender a mais de 50% dos critérios.

Tabela 1. Critérios de Qualidade

| ID | Critérios de Qualidade | Tipo |
|------|---|---------------|
| CQ1 | O estudo possui relação com a aplicação da mineração de dados para auxiliar a prever a evasão escolar do ensino básico? | Relatório |
| CQ2 | O estudo possui uma proposta (com justificativa) clara e objetiva da problemática abordada? | Relatório |
| CQ3 | O estudo realizou uma contextualização adequada do conteúdo? | Relatório |
| CQ4 | Os métodos e técnicas aplicadas foram adequadas para o objetivo da pesquisa? | Rigor |
| CQ5 | Os dados utilizados no estudo foram coletados de forma rigorosa ou recrutados de um ambiente confiável? | Rigor |
| CQ6 | O estudo realizou a análise de dados baseado em uma metodologia de análise reconhecida cientificamente? | Rigor |
| CQ7 | O estudo reportou claramente as técnicas de análise utilizadas? | Rigor |
| CQ8 | Os resultados foram claros e as conclusões justificadas? | Credibilidade |
| CQ9 | Os resultados possuem relação com o objetivo de estudo? | Credibilidade |
| CQ10 | O estudo possui qualificação igual ou acima de B2? | Credibilidade |
| CQ11 | O estudo deixou claro sua contribuição para melhoria do problema social da evasão escolar do ensino básico? | Relevância |

3.2. Condução

Essa seção expõe a segunda fase da RSL, com o objetivo de apresentar em detalhes o processo de condução. Inicialmente, a coleta dos EP considerou o período de janeiro de 2015 a fevereiro de 2021, considerando apenas EP completos. Na busca automática, um quantitativo de 271 estudos foram retornados. Já na busca manual, 1.879 trabalhos foram identificados, totalizando, ao final, 2.150 EP retornados das duas buscas. Na sequência, a etapa de seleção aconteceu por meio da leitura dos títulos e palavras-chave de cada EP identificado anteriormente. Em paralelo, eram aplicados os critérios de inclusão e exclusão, resultando ao final, 68 EP aprovados para a próxima fase. Na sequência foi realizada a segunda leitura, na qual novamente os critérios de seleção foram aplicados, mas dessa vez foram lidos título, palavras-chave, resumo, introdução, resultados e conclusões e, ao final, foram aprovados 28 estudos primários.

Na etapa final, realizou-se a terceira e última leitura, na qual os EP foram lidos na íntegra e, simultaneamente, os critérios de qualidade eram aplicados, para avaliar e identificar quais EP responderiam às PP adequadamente. Salienta-se que cada ciclo de leitura foi realizado por dois pesquisadores, tendo um suporte de decisão de dois especialistas, caso houvesse divergências. Desta forma, 21 EP formaram o escopo final desta RSL e, as evidências foram analisadas com auxílio da ferramenta de análise qualitativa ATLAS.ti. No intuito de assegurar a replicação do estudo, é disponibilizado um arquivo *online*² contendo o protocolo, o teste de sensibilidade e demais etapas da RSL.

4. Resultados e Discussões

Nesta seção são descritos os resultados e discussões das perguntas de pesquisa propostas nesta RSL, através da análise quantitativa utilizando gráficos e tabelas, como também, análise qualitativa com auxílio da ferramenta ATLAS.ti.

4.1. PP1: Quais métodos de regressão são utilizados para prever a evasão escolar no ensino básico?

Modelos de regressão utilizam modelos matemáticos treinados em dados para tomar decisões e levar à extração de informações extremamente úteis, quando aplicados a problemas complexos do mundo real. Com isso, essa questão apresenta os métodos de regressão que estão sendo utilizados na área de pesquisa, demonstrando a importância e a qualidade do seu uso para a previsão da evasão escolar no ensino básico. Então, apenas três EP fizeram uso de modelos de regressão, na qual estão apresentados na Figura 1, com suas respectivas quantidades de artigos em que cada modelo foi apresentado.

Entre os modelos supracitados, apenas regressão quantílica e regressão de vetores de suporte foram utilizados em mais de um estudo primário. Esses resultados revelam a baixa quantidade de pesquisas que estão contribuindo de forma efetiva na prática educacional, através da utilização de modelos de regressão, caracterizando novos esforços para o desenvolvimento de novas pesquisas.

Entretanto, acredita-se que a baixa adesão aos modelos de regressão, aplicados ao contexto educacional da evasão escolar do ensino básico, são decorrências do nível de complexidade da sua utilização, assim como a interpretação dos seus resultados por eles

²encurtador.com.br/wBOUX

gerados, uma vez que são modelos matemáticos aplicados para representar um problema através de uma função linear ou não linear e, com isso, os resultados obtidos são valores contínuos. Porém, os 3 estudos recrutados apresentaram que modelos de regressão utilizados estimam com maior precisão a relação entre os fatores educacionais e a evasão escolar. Também consideram como um recurso extremamente relevante no auxílio da previsão do estudante evadido e as características que influenciaram a sua evasão.

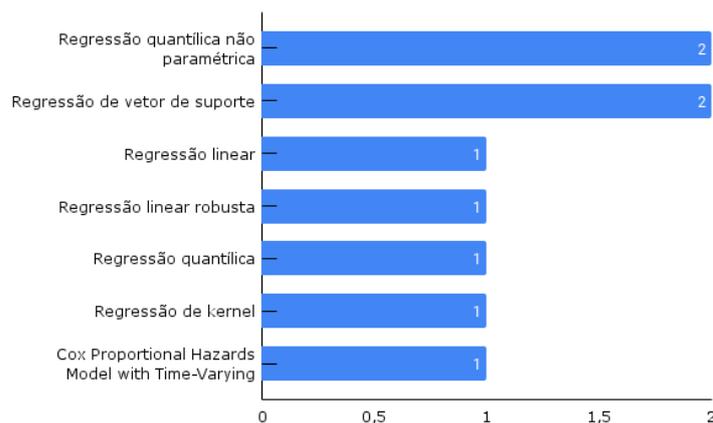


Figura 1. Modelos de regressão utilizados na previsão da evasão escolar.

4.2. PS1: Quais fatores influenciam na evasão escolar no ensino básico?

Essa questão identifica os fatores que influenciam na evasão escolar, fornecendo indicadores para o reconhecimento preliminar dos alunos que tendem a evadir do ensino básico. Como relatado anteriormente, a análise teve como suporte a ferramenta Atlas.ti. Com isso, 14 fatores com maior representatividade, considerando o nível de magnitude (quantidade de citações nos EP), seguem apresentados na Tabela 2.

Tabela 2. Fatores que influenciam na evasão escolar do ensino básico.

| Fator | Magnitude | Estudo Primário |
|---|-----------|---|
| Renda familiar de até 1 salário mínimo | 12 | EP1, EP2, EP5, EP6, EP7, EP8, EP9, EP17, EP18, EP19, EP20, EP21 |
| Faixa etária do aluno | 8 | EP1, EP2, EP7, EP8, EP9, EP11, EP12, EP13 |
| Formação e dedicação dos docentes | 6 | EP1, EP4, EP9, EP11, EP20, EP21 |
| Estudantes do sexo feminino | 4 | EP5, EP8, EP10, EP21 |
| Quantidade de alunos por turma | 3 | EP1, EP11, EP13 |
| Existência de laboratórios e computadores | 3 | EP2, EP8, EP12 |
| Turno em que o aluno estuda | 2 | EP1, EP3 |
| Localização da escola | 2 | EP2, EP5 |
| Reprovações do aluno nas disciplinas | 2 | EP8, EP9 |
| Baixa escolaridade do responsável | 2 | EP6, EP8 |
| Ausência de sala de Atendimento Educacional Especializado (AEE) na escola | 1 | EP2 |
| Aluno que trabalha mais de 4h por dia | 1 | EP17 |
| O aluno não possui computador em casa | 1 | EP8 |
| Estrutura física da escola | 1 | EP1 |

Dentre as análises que o Atlas.ti oferece, existe a possibilidade de criar redes, que representa graficamente o relacionamento entre as informações, associando aquelas que podem influenciar a existência das demais. Com isso, por questão de limitação de espaço, 8 dos 14 fatores identificados foram relacionados e estão apresentados na Figura 2.

Percebe-se, por exemplo, que os alunos de maior faixa etária se tornam propícios a evadir, considerando que existe uma pressão para trabalharem, visando sua sustentação (e familiar), interferindo na continuidade dos estudos e reprovações em disciplinas. Além disso, a estrutura física da escola possui grande influência na motivação dos alunos para continuarem os estudos. A presença de computadores é um diferencial, indicando uma melhor estrutura escolar e preocupação no ensino tecnológico.

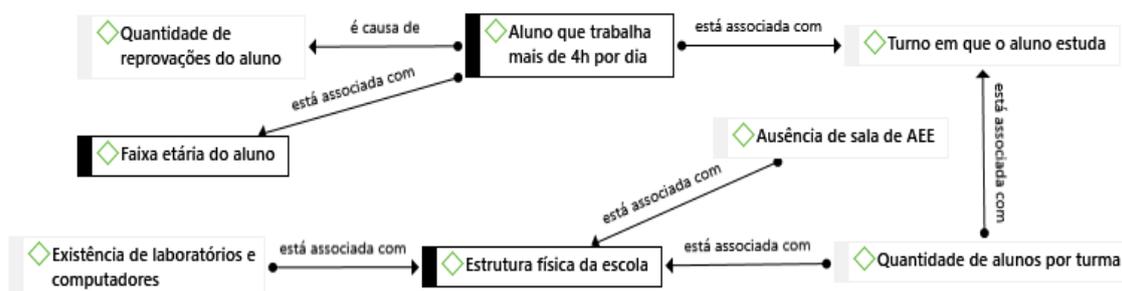


Figura 2. Associação dos fatores que influenciam na evasão escolar.

4.3. PS2: Quais ações são adotadas para evitar a evasão escolar no ensino básico?

Essa questão apresenta quais medidas já foram ou são adotadas na comunidade educacional, a fim de mitigar a evasão escolar no ensino básico, como relata a Tabela 3.

Tabela 3. Ações preventivas na redução da evasão escolar no ensino básico.

| Ação | Magnitude |
|---|-----------|
| Sistema de detecção de estudantes em risco de evasão | 5 |
| Novos estudos sobre o tema | 2 |
| Envolvimento dos pais e criação de equipes multidisciplinares | 1 |
| Oferta de cursos de artes visuais, música, teatro e dança | 1 |
| Programa Brasil Alfabetizado | 1 |
| Desenvolvimento de projetos de extensão | 1 |
| Adequação aos horários de aula de acordo com o transporte público | 1 |

Parte dos estudos sugerem o desenvolvimento e implantação de um sistema de detecção de estudantes em risco de evasão. Objetiva-se que o mesmo possa alertar os professores e pais sobre os alunos que estão potencialmente em risco de evadir. Por isso, recomenda-se o desenvolvimento de novas pesquisas sobre o tema, para compreender as causas relacionadas à evasão escolar. Assim, a equipe pedagógica, professores e demais setores da instituição, podem tomar decisões preventivas em relação à evasão escolar, de acordo com as informações extraídas de bases de dados.

Além disso, o Programa Brasil Alfabetizado³ é um exemplo de programa que tem como meta a redução da desigualdade educacional, e é voltado para a alfabetização de

³<http://portal.mec.gov.br/programa-brasil-alfabetizado>

jovens, adultos e idosos. Da mesma forma um dos EP propõe o desenvolvimento de projetos de extensão, para trabalhar com os novos alunos os conhecimentos básicos de português e matemática, por exemplo.

4.4. PS3: Quais técnicas de machine learning estão sendo utilizados para auxiliar a tomada de decisão relacionada à evasão escolar no ensino básico?

Essa pergunta identifica as técnicas de ML que estão sendo utilizadas para identificar com antecedência os estudantes favoráveis a evadir. Com isso, foi possível encontrar nos EP diferentes técnicas. Dentre elas, destacam-se os algoritmos de classificação de árvores de decisão, relatados em 10 EP, e regressão logística, aplicados em outros 7 estudos, como mostrado na Figura 2.

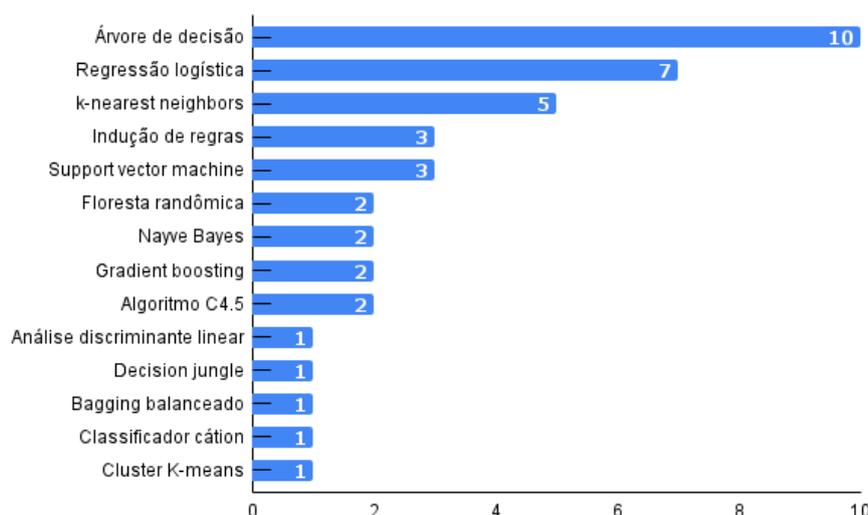


Figura 3. Técnicas de Machine Learning na previsão da evasão escolar.

5. Ameaças à Validade

Embora a RSL tenha sido realizada cuidadosamente, esta seção aborda as possíveis limitações ou ameaças que afetam a validade dos resultados, sendo elas: **(i)** a decisão sobre quais EP incluir ou excluir e como categorizá-los, pode em alguma etapa ter sido tendenciosa e, portanto, uma ameaça. De forma a minimizá-la, os processos de seleção foram realizados e revisados por quatro pesquisadores; **(ii)** a *string* de busca pode não incluir todos os termos relacionados ao tema de pesquisa. Entretanto, foram realizados testes pilotos de modo a ajustar e refinar a *string* apresentada; e **(iii)** esta RSL concentrou-se em analisar evidências específicas à evasão escolar no ensino básico, mantendo um foco específico e aumentando a ameaça de validade externa, entretanto, os resultados permitiram obter *insights* para orientar novas investigações na área de pesquisa estudada.

6. Conclusão

A educação serve como uma base sólida de uma estrutura maior e representa o indivíduo como um todo. Entender e propor soluções para os principais dilemas enfrentados pelo sistema educacional é uma forma de lidar com problemas que vivenciamos. Um dos desafios enfrentados pela sistema educacional brasileiro é a evasão escolar. Esse problema é

decorrente das várias consequências provocadas pela desigualdade social, que atinge diretamente as pessoas mais vulneráveis. Diante desse problema, as técnicas de ML tornam-se um relevante recurso para auxiliar a prever a relação entre o estudante evadido e as características que influenciaram a sua evasão. Sendo assim, esse artigo apresentou uma RSL com o objetivo de identificar as vertentes da utilização das técnicas de ML, em específico os modelos de regressão na detecção de fatores que influenciam a evasão escolar de estudantes do ensino básico.

Para esse trabalho, buscas manuais e automáticas foram conduzidas considerando o período de 2015 a 2021, e ao final 21 EP foram selecionados e as PP foram respondidas. Em seguida, com análise realizada sobre os EP, identificou-se que apenas 3 EP responderam a PP principal, apresentando como resposta a utilização de 7 modelos de regressão na previsão de evasão escolar do ensino básico. Dentre os modelos identificados destaca-se o método de regressão quantílica e regressão de vetores de suporte, pois estão presentes em 2 dos 3 EP. No entanto, evidenciou-se uma baixa quantidade de pesquisas que estão contribuindo de forma efetiva na prática educacional utilizando esses modelos, caracterizando oportunidade de trabalhos futuros.

Por conseguinte, um total de 14 fatores que influenciam na evasão escolar foram identificados, com destaque aos estudantes com renda familiar inferior a 1 salário mínimo presente em 12 EP e a faixa etária presente em 8 dos 21 EP. Também ficou evidenciável como respostas para PS1 que a formação e dedicação dos docentes são fatores influentes da evasão escolar, pois encontram-se presentes em 6 EP. Diante disso, conclui-se que as evidências fornecem os indicadores que ajudarão no reconhecimento preliminar dos estudantes que tendem a evadir do ensino básico e, com isso, pode-se abrir rupturas para o desenvolvimento e aplicabilidade de políticas públicas na necessidade de evitar a evasão.

Posteriormente, foram analisadas quais ações que já foram ou estão sendo adotadas na comunidade educacional, para mitigar a evasão escolar no ensino básico. Como resposta para PS2, 5 EP ressaltam a ideia de desenvolver um sistema de detecção de estudantes em risco de evasão, em complementar os outros EP relataram ações como novas pesquisas, desenvolvimento de projetos extensionistas e a união entre escola, pais e sociedade. Portanto, os métodos de ML que estão sendo utilizados na área de pesquisa, e foram apresentados como resposta da PS3, seus resultados identificaram a alta utilização dos modelos de classificação presentes nos EP. Destacando-se as árvores de decisão presentes em 10 EP e regressão logística utilizada por 7 estudos.

Por fim, acredita-se que os resultados apresentados proporcionaram alcançar os objetivos da revisão sistemática da literatura e, conseqüentemente, permitiram a identificação de algumas lacunas. Portanto, como trabalhos futuros, propõe-se o desenvolvimento de um sistema baseado na previsão dos fatores relacionados à evasão escolar no ensino básico, como também, a utilização de modelos baseados em ML identificados neste estudo e que se apresentam como melhores técnicas no desempenho da previsão ou identificação dos estudantes propensos a evadir.

Referências

- Adelman, M., Haimovich, F., Ham, A., and Vazquez, E. (2018). Predicting school dropout with administrative data.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

- Calixto, K., Segundo, C., and de Gusmão, R. P. (2017). Mineração de dados aplicada a educação: um estudo comparativo acerca das características que influenciam a evasão escolar. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 1447.
- Colpo, M. P., Primo, T. T., Pernas, A. M., and Cechinel, C. (2020). Mineração de dados educacionais na previsão de evasão: uma rsl sob a perspectiva do congresso brasileiro de informática na educação. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 1102–1111. SBC.
- do Desenvolvimento Humano, P. R. (2019). Além do rendimento, além das médias, além do presente: Desigualdades no desenvolvimento humano no século xxi. *Programa das Nações Unidas para o Desenvolvimento*.
- do Nascimento, R. L. S., da Cruz Junior, G. G., and de Araújo Fagundes, R. A. (2018). Mineração de dados educacionais: Um estudo sobre indicadores da educação em bases de dados do inep. *RENOTE*, 16(1).
- Dyba, T., Dingsoyr, T., and Hanssen, G. K. (2007). Applying systematic reviews to diverse study types: An experience report. In *First international symposium on empirical software engineering and measurement (ESEM 2007)*, pages 225–234. IEEE.
- Júnior, O. d. G. F., Rodrigues, W. R. M., Barbirato, J. C. C., and de Barros Costa, E. (2019). Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: um estudo de caso em escolas municipais de maceió. *RENOTE*.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.
- Lima, J. V., Júnior, M. d. M. A., Moya, A., Almeida, R., Anjos, P., Lencastre, M., Fagundes, R. A. d. A. F., and Alencar, F. (2019). As metodologias ativas e o ensino em engenharia de software: uma revisão sistemática da literatura. In *Anais do Workshop de Informática na Escola*, volume 25, pages 1014–1023.
- Lopes Filho, J. A. B. and Silveira, I. F. (2021). Detecção precoce de estudantes em risco de evasão usando dados administrativos e aprendizagem de máquina. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E40):480–495.
- Marques, L. T., De Castro, A. F., Marques, B. T., Silva, J. C. P., and Queiroz, P. G. G. (2019). Mineração de dados auxiliando na descoberta das causas da evasão escolar: Um mapeamento sistemático da literatura. *RENOTE*, 17(3):194–203.
- OECD (2015). Programme for international student assessment (pisa) results from pisa.
- SEDU, G. D. E. D. E. et al. (2018). Diretrizes para a prevenção do abandono e da evasão escolares no ensino fundamental anos finais e no ensino médio. *SEDU - Secretaria da Educação*.
- UNICEF et al. (2020). Unicef e ações de educação em tempos de pandemia. *Brasília: UNICEF*.